# Digging, Reaching, and Learning: An Update on the First Year of the HathiTrust Research Center's Librarian Training Program

Eleanor Dickson, Leanne Nay, Nicholae Cline, Harriett Green

www.teach.htrc.illinois.edu

Digging Deeper, Reaching Further: Libraries Empowering Users to Mine the HTDL Resources

# PROJECT GOALS

This 3-year project aims to:

- Arm academic librarians with practical skills in text mining methods and knowledge of trends in digital scholarship.
- Situate the academic library as a space for encountering and engaging with "big data" tools and methodologies vis-à-vis the HTRC.
- Train librarians in digital scholarship skills in a way that recognizes the decentralized and fluid nature of digital humanities research support in libraries.

### 2017-2018 ITERATIVE WORKSHOP MODEL Workshop Roadshow To be held at regional Spring 2016 Spring 2016 and national conferences Held at Illinois and Will be held at all Assess & Indiana partner institutions curriculum For librarians and Incorporating Assess & Fall 2016 Open Assess & LIS students feedback Educational curriculum from the fall Resources Will be held at all partner institutions Will be shared on our website for others to Using revised activities reuse

# **PROJECT PARTNERS**

- University of Illinois
- Indiana University
- Lafayette College
- Northwestern University
- University of North Carolina at Chapel Hill

## THE HTRC

The HathiTrust Research Center is the research arm of the HathiTrust. It facilitates scholarly research using the HathiTrust Digital Library by providing mechanisms for researchers to study HathiTrust content using computational tools for text analysis.

# RESEARCH CENTER

## INSTRUCTIONAL DESIGN

# Outcomeoriented

- Learning outcomes written for each module
- Emphasize concepts

# Modular

- Scaffolded instruction
- Translatable into OER
- Partner institutions can remix and reorganize content

Applicable to distinct audiences & unique lab set-ups

Reusable

In pilot workshops, test scenarios

- Hands-on
- Using HTRC web-based tools
- Basic command line and Python skills
- Building skills in OpenRefine

## CURRICULUM

## **Module 1: Introduction**

Intro to text mining and to the HathiTrust

## **Module 2: Gathering Textual Data**

- Two lessons, one for beginners and one more advanced
- How to create textual datasets/HTRC worksets
- Hands-on: build a collection of text to analyze

## Module 3: Managing Textual Data

- Intro to data curation
- How to prepare a textual dataset for analysis
- Hands-on: run script to remove HTML tags, or clean spreadsheet in OpenRefine

## **Module 4: Performing Text Analysis**

- Two lessons, one for beginners and one more advanced
- How to run a text analysis algorithm
- Hands-on: run algorithm in HTRC portal, or run a script to study HTRC Extracted Features files

## Module 5: Visualizing Text Data with HT+BW

- How to use HathiTrust + Bookworm to visualize HTDL data and metadata
- Hands-on: track trends in how words were used over time

# ASSESSMENT

- Administer assessment survey after each workshop
- IRB-approved with goal of sharing results in the future

## Sample questions:

- What was the most interesting thing you learned?
- What could be improved or added?
- How effective was this workshop in teaching you about the HTRC?
- How do you see yourself applying what you learned?